Complementary proofs to "Approximation of discrete BSDE using least-squares regression"

E. Gobet[a,1,*], P. Turkedjiev[b]

[a] *Ecole Polytechnique and CNRS, Centre de Mathématiques Appliquées, Route de Saclay, 91128 Palaiseau, France*
[b] *Humboldt Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, 10099 Berlin, Germany*

Abstract

This short note gives complementary proofs to our work [GT11], of which we follow the notations and assumptions.

1. Assumption ($\mathbf{A_F}$-iii) for non-uniform grids

The time grid $(t_k = T - T(1-k/N)^{1/\theta_\pi})_{0\leq k\leq N}$ with $\theta_\pi \in (0,1]$ satisfies

$$C_\pi := \sup_{k<N} \frac{\Delta_k}{(T-t_k)^{1-\theta_L}} \leq \frac{T^{\theta_L}}{\theta_\pi} \frac{1}{N^{1\wedge \frac{\theta_L}{\theta_\pi}}},$$
$$R_\pi := \sup_{0\leq k\leq N-2} \frac{\Delta_k}{\Delta_{k+1}} \leq \frac{1}{\theta_\pi}\left(1 \vee \left(\frac{1}{2\theta_\pi}\right)^{\frac{1}{\theta_\pi}-1}\right),$$

where $\Delta_k = t_{k+1} - t_k$ and $\theta_L \in (0,1]$.

Proof. Set $1/\theta_\pi = \mu \geq 1$ and $g(x) = 1-(1-x)^\mu$: we have $t_k = Tg(k/N)$. Note that g is increasing and concave; thus we have

$$\frac{\Delta_k}{(T-t_k)^{1-\theta_L}} \leq \frac{\frac{\mu T}{N}(1-k/N)^{\mu-1}}{T^{1-\theta_L}(1-k/N)^{\mu(1-\theta_L)}} = \frac{T^{\theta_L}}{\theta_\pi N}(1-k/N)^{\theta_L/\theta_\pi - 1}$$

and the bound on C_π follows by considering either $\theta_L \geq \theta_\pi$ or $\theta_L < \theta_\pi$.

Now, we study R_π. Since g is concave, we have $\Delta_{k-1} \geq \Delta_k \geq \cdots \geq \Delta_{N-1} = TN^{-\mu}$ and $\Delta_{k-1} \leq \frac{\mu T}{N}(1-k/N)^{\mu-1}$. This gives a first upper bound for the

*Corresponding author
Email addresses: emmanuel.gobet@polytechnique.edu (E. Gobet), turkedji@math.hu-berlin.de (P. Turkedjiev)

[1] This research is part of the Chair *Financial Risks* of the *Risk Foundation*, the Chair *Derivatives of the Future* and the Chair *Finance and Sustainable Development*.

n_0-last times $k = N - n_0, \dots, N-1$ (with $n_0 \geq 1$):

$$\frac{\Delta_{k-1}}{\Delta_k} \leq \frac{\Delta_{k-1}}{\Delta_{N-1}} \leq \frac{\frac{\mu T}{N}(1-k/N)^{\mu-1}}{TN^{-\mu}} = \mu(N-k)^{\mu-1} \leq \mu n_0^{\mu-1}. \tag{1}$$

We are now in a position to complete the upper bound on R_π.

- $\mu \in [1,2]$: we prove $\frac{\Delta_{k-1}}{\Delta_k} \leq \mu$. For $k = N-1$, the inequality is true owing to (1). Now take $k < N-1$. Since g'' is non-increasing ($\mu \in [1,2]$), we have $\Delta_k \geq \frac{\mu T}{N}(1-k/N)^{\mu-1} + \frac{T}{2N^2} g''((k+1)/N)$, and we easily deduce

$$\frac{\Delta_{k-1}}{\Delta_k} \leq \frac{\frac{\mu T}{N}(1-k/N)^{\mu-1}}{\frac{\mu T}{N}(1-k/N)^{\mu-1} - \frac{T}{2N^2}\mu(\mu-1)(1-(k+1)/N)^{\mu-2}} \leq \frac{1}{1-\frac{(\mu-1)}{2}} \leq \mu.$$

- $\mu \geq 2$: we prove $\frac{\Delta_{k-1}}{\Delta_k} \leq \mu\big(\frac{\mu}{2}\big)^{\mu-1}$. Set $n_0 = \lfloor \frac{\mu}{2} \rfloor$: $n_0 \leq \frac{\mu}{2} < n_0 + 1$. For $k \geq N - n_0$, the announced upper bound directly follows from (1). Now take $k \leq N - n_0 - 1$ (which implies $N - k > \frac{\mu}{2}$): g'' being non-decreasing for $\mu \geq 2$, we have

$$\begin{aligned}
\Delta_k &\geq \frac{\mu T}{N}(1-k/N)^{\mu-1} + \frac{T}{2N^2} g''(k/N) \\
&= \frac{\mu T}{N}(1-k/N)^{\mu-1}\big[1 - \frac{(\mu-1)}{2}(N-k)^{-1}\big] \\
&> \frac{\mu T}{N}(1-k/N)^{\mu-1}\frac{1}{\mu} \geq \Delta_{k-1}\frac{1}{\mu}.
\end{aligned}$$

$\square$

2. Proof of Proposition 3.3.

Proposition 3.3. Assume $(\mathbf{A}'_\xi\textbf{-i})$ and $(\mathbf{A_F})$. For any $R \in [0, +\infty]$ and for any π with N large enough (such that $C_\pi L_f^2 \leq \frac{1}{12q}$), the following almost sure error bounds on $Y_i - Y_i^R$ and $Z_i - Z_i^R$ hold for any $0 \leq i < N$:

$$|Y_i - Y_i^R| \leq C_y \exp\big(\frac{T}{8} + \frac{12qL_f^2}{\theta_L}T^{\theta_L}\big)\exp\big(-\frac{1}{4}R^2\big)\sqrt{N},$$

$$\big(\sum_{k=i}^{N-1} \mathbb{E}_i|Z_k - Z_k^R|^2\Delta_k\big)^{\frac{1}{2}} \leq C_y \exp\big(\frac{12qL_f^2}{\theta_L}T^{\theta_L}\big)\big(8q + T\exp(\frac{T}{4})\big)^{\frac{1}{2}}\exp\big(-\frac{1}{4}R^2\big)\sqrt{N}.$$

Proof. We set $\mathcal{T}_R := \mathbb{E}([\mathcal{N} - (-R) \vee \mathcal{N} \wedge R]^2)$ where $\mathcal{N}$ is a Gaussian random variable with mean 0 and variance 1. An explicit computation gives

$$\mathcal{T}_R = 2\big(\mathbb{P}(\mathcal{N} > R)(R^2+1) - R\frac{e^{-\frac{1}{2}R^2}}{\sqrt{2\pi}}\big) \leq 2\mathbb{P}(\mathcal{N} > R)(R^2 + 1 - R^2) \leq 2e^{-\frac{1}{2}R^2},$$

where the two last inequalities are derived from the Mill inequality and the Markov exponential inequality.

Now, we follow the arguments of Lemma 3.1 and we consider $\gamma \in (0,+\infty)^N$ such that $8q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1$ for $0 \leq k < N$. Define $\Delta Y_k := Y_k - Y_k^R$ and $\Delta Z_k := Z_k - Z_k^R$.

Preliminary bound on ΔZ. Applying the Cauchy-Schwartz inequality and the almost sure bounds on Y and Y^R (Proposition 3.2), we obtain:

$$\begin{aligned}
\Delta_k|\Delta Z_k|^2 &= \Delta_k^{-1}\big|\mathbb{E}_k\big[Y_{k+1}\Delta W_k - Y_{k+1}^R[\Delta W_k]_w\big]\big|^2 \\
&\leq 2\Delta_k^{-1}\big|\mathbb{E}_k[Y_{k+1}(\Delta W_k - [\Delta W_k]_w)]\big|^2 + 2\Delta_k^{-1}\big|\mathbb{E}_k\big[\Delta Y_{k+1}[\Delta W_k]_w\big]\big|^2 \\
&\leq 2qC_y^2\mathcal{T}_R + 2q\big(\mathbb{E}_k[\Delta Y_{k+1}^2] - (\mathbb{E}_k[\Delta Y_{k+1}])^2\big). \qquad (2)
\end{aligned}$$

Bound on ΔY. Using Young's inequality $(a+b)^2 \leq (1+\Delta_k\gamma_k)a^2 + (1+\frac{1}{\Delta_k\gamma_k})b^2$, the Lipschitz property of $(y,z) \mapsto f_k(y,z)$, and using (2), we obtain

$$\begin{aligned}
\Delta Y_k^2 \leq& (1+\Delta_k\gamma_k)(\mathbb{E}_k[\Delta Y_{k+1}])^2 \\
&+ 2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\Delta_k(\mathbb{E}_k[\Delta Y_{k+1}^2] + |\Delta Z_k|^2) \qquad (3) \\
\leq& \big(1+\Delta_k\gamma_k - 4q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\big)(\mathbb{E}_k[\Delta Y_{k+1}])^2 \qquad (4) \\
&+ 2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}(\Delta_k + 2q)\mathbb{E}_k[\Delta Y_{k+1}^2] \\
&+ 4qC_y^2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\mathcal{T}_R.
\end{aligned}$$

The condition $8q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \leq 1$ ensures that $1 + \Delta_k\gamma_k - 4q(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}} \geq 0$; this given, we may use Jensen's inequality on the term (4) to obtain:

$$\begin{aligned}
\Delta Y_k^2 &\leq \big(1+\Delta_k\gamma_k + 2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\Delta_k\big)\mathbb{E}_k[\Delta Y_{k+1}^2] \\
&\quad + 4qC_y^2(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\mathcal{T}_R \\
&\leq \big(1+\Delta_k\gamma_k + \frac{\Delta_k}{4}\big)\mathbb{E}_k[\Delta Y_{k+1}^2] + \frac{1}{2}C_y^2\mathcal{T}_R
\end{aligned}$$

using again the relation between Δ_k and $1/\gamma_k$. Multiplying by $\lambda_k := \prod_{j=0}^{k-1}(1 + \Delta_j\gamma_j + \frac{\Delta_j}{4})$, taking conditional expectation $\mathbb{E}_i$, and summing over $k = i,\dots,N-1$, we obtain a pointwise uniform bound for ΔY_i^2:

$$\Delta Y_i^2\Gamma_i \leq \Delta Y_i^2\lambda_i \leq \frac{1}{2}C_y^2 e^{T/4}\Gamma_N N\mathcal{T}_R. \qquad (5)$$

Final bound on ΔZ. (2) yields:

$$\sum_{k=i}^{N-1} \Gamma_k \mathbb{E}_i[|\Delta Z_k|^2]\Delta_k \le 2qC_y^2\Gamma_N N\mathcal{T}_R + 2q\sum_{k=i}^{N-1}\big(\mathbb{E}_i[\Delta Y_{k+1}^2] - \mathbb{E}_i(\mathbb{E}_k[\Delta Y_{k+1}])^2\big)\Gamma_{k+1}$$
$$\le 2qC_y^2\Gamma_N N\mathcal{T}_R + 2q\sum_{k=i}^{N-1}\big(\mathbb{E}_i[\Delta Y_k^2] - (1+\Delta_k\gamma_k)\mathbb{E}_i(\mathbb{E}_k[\Delta Y_{k+1}])^2\big)\Gamma_k.$$

Substituting the inequality (3), we obtain

$$\sum_{k=i}^{N-1} \Gamma_k \mathbb{E}_i[|\Delta Z_k|^2]\Delta_k$$
$$\le 2qC_y^2\Gamma_N N\mathcal{T}_R + 4q\sum_{k=i}^{N-1}(\Delta_k + \frac{1}{\gamma_k})\frac{L_f^2}{(T-t_k)^{1-\theta_L}}\Delta_k\Gamma_k(\mathbb{E}_i[\Delta Y_k^2] + \mathbb{E}_i[|\Delta Z_k|^2])$$
$$\le 2qC_y^2\Gamma_N N\mathcal{T}_R + \frac{1}{2}\sum_{k=i}^{N-1}\Delta_k\Gamma_k(\mathbb{E}_i[\Delta Y_k^2] + \mathbb{E}_i[|\Delta Z_k|^2])$$

taking into account the relation between π and γ. Thus, we have

$$\sum_{k=i}^{N-1} \Gamma_k \mathbb{E}_i[|\Delta Z_k|^2]\Delta_k \le C_y^2\Gamma_N N\mathcal{T}_R\big(4q + \frac{T}{2}\exp(\frac{T}{4})\big). \tag{6}$$

Observe that $\gamma_k := \frac{24qL_f^2}{(T-t_k)^{1-\theta_L}}$ defines an admissible choice, provided that $C_\pi L_f^2 \le \frac{1}{12q}$. It gives $1 \le \Gamma_i \le \Gamma_N \le \exp\big(\frac{24qL_f^2}{\theta_L}T^{\theta_L}\big)$. Plugging this estimate into (5) and (6), and using the bound on $\mathcal{T}_R$, we obtain the final result.
□

3. Proof of Lemma 4.2

Lemma 4.2. With the current notation and assumptions, for all m we have

$$\bar{y}_k^{R,M}(X_k^m) = \mathbb{E}_N^M\big[\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i\big],$$

$$\Delta_k\bar{z}_{l,k}^{R,M}(X_k^m) = \mathbb{E}_N^M\big[[\Delta\tilde{W}_{l,k}^m]_w\big(\Phi(\tilde{X}_N^{k,m}) + \sum_{i=k+1}^{N-1} f_i(\tilde{X}_i^{k,m}, y_{i+1}^{R,M}(\tilde{X}_{i+1}^{k,m}), z_i^{R,M}(\tilde{X}_i^{k,m}))\Delta_i\big)\big].$$

Proof. We start with a standard result. Let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$, such that $\mathcal{G} \perp\!\!\!\perp \mathcal{H}$. Let $F : \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ be bounded and $\mathcal{G}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable, and $U : \Omega \to \mathbb{R}^d$ be $\mathcal{H}$-measurable. Then, by the Monotone Class Theorem for

functions, $\mathbb{E}[F(U)|\mathcal{H}] = j(U)$ where $j(h) = \mathbb{E}[F(h)]$ for all $h \in \mathbb{R}^d$.
In order to apply the above result, we require some standard results about the ghost path $(\tilde{X}, \Delta\tilde{W})$. Let k be fixed. Since $\tilde{X}$ is a Markov chain, then for all $i > k$ there is a mapping $V_i : \Omega\times\mathbb{R}^d \to \mathbb{R}^d$ measurable with respect to $\mathcal{G}_i\otimes\mathcal{B}(\mathbb{R}^d)$ such that $\tilde{X}_i^{x,k} = V_i(x)$, where the filtration $(\mathcal{G}_i)_{k<i\leq N}$ is independent of $\mathcal{F}_N^M$ and $\Delta\tilde{W}_k^k$ is $\mathcal{G}_{k+1}$-measurable.
Now, by defining

$$\begin{cases} F_1(x) & := \Psi_k^{R,M}(x, V_{k+1}(x), \dots, V_N(x)), \\ F_2(x) & := [\Delta\tilde{W}_k^{k,m}]_w \Psi_{k+1}^{R,M}(V_{k+1}(x), V_{k+2}(x), \dots, V_N(x)), \end{cases}$$

the result of the previous paragraph can be applied, because F_1 and F_2 are $\mathcal{G}_N \otimes \mathcal{B}(\mathbb{R}^d)$-measurable, hence the representations for $\bar{y}_k^{R,M}(X_k^m)$ and $\bar{z}_k^{R,M}(X_k^m)$. $\square$

4. Proof of Lemma 1 in Appendix B

Lemma 1. Let $\mathcal{G}$ be a countable set of functions $g : \mathbb{R}^d \mapsto [0, B]$ with $B > 0$. Let $X, X^1, \dots, X^M$ $(M \geq 1)$ be i.i.d. $\mathbb{R}^d$ valued random variables. For any $\alpha > 0$ and $\varepsilon \in (0, 1)$ one has

$$\mathbb{P}\big(\sup_{g\in\mathcal{G}} \frac{\frac{1}{M}\sum_{m=1}^M g(X^m) - \mathbb{E}[g(X)]}{\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \mathbb{E}[g(X)]} > \varepsilon\big) \leq 4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\alpha\varepsilon}{5}, \mathcal{G}, X^{1:M}\big)\big) \exp\big(-\frac{3\varepsilon^2\alpha M}{40B}\big),$$

$$\mathbb{P}\big(\sup_{g\in\mathcal{G}} \frac{\mathbb{E}[g(X)] - \frac{1}{M}\sum_{m=1}^M g(X^m)}{\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \mathbb{E}[g(X)]} > \varepsilon\big) \leq 4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\alpha\varepsilon}{8}, \mathcal{G}, X^{1:M}\big)\big) \exp\big(-\frac{6\varepsilon^2\alpha M}{169B}\big).$$

Proof. The first inequality is stated in [GKKW02, Theorem 11.6] for $B \geq 1$. For $B \in (0, 1)$, we rescale the class of functions $\{g/B : g \in \mathcal{G}\}$ (now bounded by 1), replace α by α/B and apply the previous case: this gives the announced upper bound.
To establish the second inequality, we adapt the proof of the first inequality from the proof of [GKKW02, Theorem 11.6]. The first step consists in taking a ghost sample $\tilde{X}^{1:M}$ and observing that for a given $g \in \mathcal{G}$, $\mathbb{E}[g(X)] - \frac{1}{M}\sum_{m=1}^M g(X^m) > \varepsilon\big(\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \mathbb{E}[g(X)]\big)$ and $\mathbb{E}[g(X)] - \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) \leq \frac{\varepsilon}{4}\big(\alpha + \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) + \mathbb{E}[g(X)]\big)$ imply

$$\begin{aligned}(1 + \frac{5\varepsilon}{8})\big(\frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m) - \frac{1}{M}\sum_{m=1}^M g(X^m)\big) \\ > \frac{3\varepsilon}{8}\big(2\alpha + \frac{1}{M}\sum_{m=1}^M g(X^m) + \frac{1}{M}\sum_{m=1}^M g(\tilde{X}^m)\big) + \frac{3\varepsilon}{4}\mathbb{E}[g(X)].\end{aligned}$$

Since the r.h.s. positive, the l.h.s. is also positive; using $\frac{13}{8} \geq 1 + \frac{5\varepsilon}{8}$ implies

$$\frac{1}{M}\sum_{m=1}^{M} g(\tilde{X}^m) - \frac{1}{M}\sum_{m=1}^{M} g(X^m) > \frac{3\varepsilon}{13}\big(2\alpha + \frac{1}{M}\sum_{m=1}^{M} g(X^m) + \frac{1}{M}\sum_{m=1}^{M} g(\tilde{X}^m)\big).$$

Then we proceed as in [GKKW02, pp. 205-207] to show that the probability to estimate is bounded by

$$2\mathbb{P}\Big(\exists g \in \mathcal{G} : \frac{1}{M}\sum_{m=1}^{M} g(\tilde{X}^m) - \frac{1}{M}\sum_{m=1}^{M} g(X^m) > \frac{3\varepsilon}{13}\big(2\alpha + \frac{1}{M}\sum_{m=1}^{M} g(X^m) + \frac{1}{M}\sum_{m=1}^{M} g(\tilde{X}^m)\big)\Big)$$

for $M > \frac{8B}{\varepsilon^2\alpha}$ (however for $M \leq \frac{8B}{\varepsilon^2\alpha}$ the upper bound in Lemma 1 is obviously true). The rest of the proof is identical to [GKKW02, pp. 208-210], except that one should take a $\mathbf{L}_1$ δ-cover of $\mathcal{G}$ w.r.t. $X^{1:M}$ with $\delta = \frac{\alpha\varepsilon}{8}$ (instead of $\delta = \frac{\alpha\varepsilon}{5}$). It leads to a new upper bound, $4\mathbb{E}\big(\mathcal{N}_1\big(\frac{\alpha\varepsilon}{8}, \mathcal{G}, X^{1:M}\big)\big)\exp\big(-\frac{6\varepsilon^2\alpha M}{169B}\big)$. $\square$

References

[GKKW02] L. Gyorfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics, 2002.

[GT11] E. Gobet and P. Turkedjiev. Approximation of discrete BSDE using least-squares regression. *Submitted*, 2011.